Optimising Recombinant Fusion Protein Production in *Pichia pastoris:* Sequence modifications and a simple method for the generation of multi-copy strains

Prashant Pyati[1], Elaine Fitches[2] and John A. Gatehouse*[1]

1. School of Biological and Biomedical Sciences, Durham University, South Road, Durham DH1 3LE, UK

2. The Food and Environment Research Agency, Sand Hutton, York, YO41 1LZ, UK.

* Corresponding author:
E-mail J.A.Gatehouse@durham.ac.uk
Phone +44 191 334 1264
FAX +44 191 334 1201

Abstract

A recombinant ~~construct encoding for the expression of an insecticidal~~ fusion protein ~~comprised of~~containing ~~the spider venom peptide~~ ω-hexatoxin-Hv1a (Hv1a), from the venom of the funnel web spider *Hadronyche versuta*, linked to ~~the~~ snowdrop lectin (*Galanthus nivalis* agglutinin; ~~(~~GNA) has been shown to be an effective oral insecticide against lepidopteran larve. The construct for producing the Hv1a/GNA fusion protein in *P. pastoris* has been modified ~~to improve levels of intact recombinant protein recoverable from fermented culture supernatants.~~ by s~~S~~ite directed mutagenesis to remove a potential Kex2 cleavage site at the C-terminus of the Hv1a peptide. The modifed construct resulted in ~~markedly~~ increased levels of intact fusion protein expressed in wild type, but not protease deficient, *P. pastoris* strains, improving levels of intact recombinant protein recoverable from culture supernatants. ~~indicative of increased resistance to yeast proteases.~~ Injection assays of purified fusion protein in lepidopteran larvae demonstrated that sequence modification did not affect the insecticidal activity of the recombinant toxin. The incorporation of a C-terminal (histidine)$_6$ tag in the modified construct enabled a single step purification of the fusion protein from fermenter supernatants ~~derived from bench-top fermentation~~. A straightforward method for producing multicopy expression plasmids for production of recombinant proteins in *P. pastoris* is described, which does not rely on multiple integrations to give clones of *P. pastoris* containing high copy numbers of the introduced gene. The method has been used to increase production of the secreted recombinant fusion protein in a pilot scale laboratory fermentation system by almost 10-fold on a per litre of culture basis, providing a means to allow evaluation as a commercial biopesticide.

Keywords: fusion protein; *Pichia pastoris*; Kex 2; cleavage; protease deficient; wild type; multi-copy

Formatted: Font: Italic

Formatted: Subscript

Introduction

The methylotrophic yeast *Pichia pastoris* has established itself as the eukaryotic expression system of choice for large-scale production of recombinant proteins, which do not fold well in prokaryotes. The system has a major advantage over *Saccharomyces* spp. in that it contains an efficient secretion system, which can direct large amounts of recombinant protein into the culture medium, making purification of the desired product straightforward. Although its limited capacity for protein glycosylation limits its use for production of mammalian glycoproteins, it is an ideal host for producing small proteins with a high content of disulphide bridges, which require the conditions of a eukaryotic endoplasmic reticulum compartment to fold to biologically active forms ([1] Daly & Hearn, 2005). Typical of this use is the production of snowdrop lectin (*Galanthus nivalis* agglutinin, GNA), which is secreted as a fully active folded protein (Raemaekers et al 1999 [2] Baumgartner *et al.*, 2003[3]), whereas expression in *E. coli* results in an insoluble inclusion body, from which GNA can only be recovered by a process of denaturation and renaturation ([4] Sulan et. al., 2005).

Comment [JG1]: Might as well quote our own paper here - **Production and purification of active snowdrop lectin in Escherichia coli** **M Longstaff, K S POWELL, J A Gatehouse, R Raemaekers, C A Nevell, and WDO Hamilton** *European Journal Of Biochemistry*, 252(1), pp.59–65.

Snowdrop lectin (*Galanthus nivalis* agglutinin; GNA) has been used as a "carrier" domain in the production of recombinant insecticidal fusion proteins, where it confers oral toxicity on peptide toxins from scorpions and spiders ([5-9] Fitches *et al.*, 2004; 2010; 2012; Down *et al.*, 2006; Trung *et al.*, 2006). Oral toxicity is achieved via delivery of attached toxins, by GNA, to the central nervous system of target pests. These toxins themselves contain multiple disulphide bonds, and expression in *E. coli* gives products that require careful refolding to give any activity; in contrast, fusion proteins where the toxin has full biological activity can be produced directly in *P. pastoris*. As a result, production of fusion proteins on a scale to allow them to be used as pesticides has become possible, but is limited by the economic cost of the fermentation system. To decrease costs, and generate an economically viable product, yields of the recombinant fusion protein must be increased as far as possible. The recombinant insecticidal fusion protein Hv1a/GNA, containing ω-hexatoxin-Hv1a (Hv1a), from the venom of the funnel web spider *Hadronyche versuta*, linked to GNA has previously been shown to be orally active against lepidopteran larvae, and has been selected as a "best candidate" in terms of potential for development as a biopesticide ([7] Fitches et. al., 2012).

To maximise expression of recombinant proteins in *P. pastoris*, investigators have often used constructs driven by the alcohol oxidase (*AOX1*) promoter, where expression can be induced by addition of methanol to a growing culture. Experiments with fusion protein production have shown that better expression was achieved with the constitutively expressed *GAP* promoter, which is also advantageous for industrial production in not requiring a methanol feed.

Previous studies have shown that insecticidal fusion proteins are prone to degradation by yeast extracellular proteases during production by bench-top fermentation (Fitches [5,7,9] *et al.*, 2004; 2012; Trung *et al.*, 2006). Proteolysis occurs predominantly at or near the linker region between the insecticidal peptide and the carrier protein resulting in a reduction in yields of intact protein. Proteolysis is particularly evident when fusion proteins are expressed using wild type X33 *P. pastoris* strain. The use of the *P. pastoris* strain SMD1168H, which is deficient in the extracellular vacuole peptidase A (pep4), responsible for activating carboxypeptidase Y and protease B1, has been found to reduce proteolysis allowing for an increase in yields of intact fusion protein to be achieved. However, X33 is the strain of choice for large-scale production, as recombinant protein production levels in the protease deficient strain tend to be lower than the more robust wild type strain ([10] Gilleson et al. 1998). In this study the removal of a potential Kex2 cleavage site present in the Hv1a toxin sequence has been shown to significantly reduce proteolysis in the wild type strain. A comparison of levels of intact fusion protein obtained by bench-top fermentation using X33 and protease deficient strains have shown that yields from X33 cells can be almost double that obtained using protease deficient strain.

Previous work has also shown that engineering multiple copies of an expression construct into the *Pichia* genome can result in increases in expression levels ([11,12] Vassileva *et. al.*, 2001, Mansur *et. al.*, 2005); with recombinant protein production increasing with increasing copy number to an optimum, after which further increasing the copy number results in decreased expression ([13] Zhu *et. al.*, 2009). The experiments described in this paper show that the multicopy strategy for maximising recombinant protein expression is also applicable to constructs based on a constitutive promoter, and identify an optimum transgene copy number for engineered *P. pastoris* strains producing the insecticidal Hv1a/GNA fusion protein.

Materials and Methods

Materials and recombinant techniques

General molecular biology protocols were as described in Sambrook & Russell ([14] 2001) except where otherwise noted. Subcloning was carried out using the TOPO cloning kit (pCR2.1 TOPO vector; Invitrogen). *Pichia pastoris* wild type X33 and SMD1168H (protease A deficient) strains, the expression vector pGAPZαB, and Easycomp *Pichia* transformation kit were from Invitrogen. Gel extraction was carried out using Qiagen gel extraction kits. Oligonucleotide primers were synthesised by Sigma-Genosys Ltd. and restriction endonucleases were purchased from Fermentas. Plasmid DNA was prepared using Promega Wizard miniprepkits. T4 polynucleotide kinase and T4 DNA ligase were supplied by Promega. Phusion polymerase was from New England BioLabs. GNA was produced as a recombinant protein in yeast using a

clone generated as previously described ([13] Raemaekers *et al.*, 1999). Polyclonal anti-GNA antibodies (raised in rabbits) were prepared by Genosys Biotechnologies, Cambridge, UK. Chemiluminescence detection reagents (coumaric acid and luminol) were supplied by Sigma.

All DNA sequencing was carried out using dideoxynucleotide chain termination protocols on Applied Biosystems automated DNA sequencers by the DNA Sequencing Service, School of Biological and Biomedical Sciences, University of Durham, UK. Sequences were checked and assembled using Sequencher software running on Mac OS computers.

Construct preparation

The generation of a construct encoding for the mature omega peptide (Hv1a) linked to the N-terminus of GNA has been previously reported ([7] Fitches *et al.*, 2012). A potential Kex2 cleavage site present in the Hv1a peptide sequence was removed and a construct was created whereby residue number 34 in the toxin (lysine; K) was replaced by a glutamine (Q) by site directed mutagenesis (Fig. 1a). The Hv1a sequence was modified by PCR using primers encoding a 5' *Pst* I site and a 3' primer encoding a modified C-terminus (as above) and *Not* I site. The PCR product was restricted and ligated into similarly digested Hv1a/GNApGAPZαB to create MODHv1a/GNApGAPZαB. A histidine tag was subsequently incorporated at the C-terminus of the fusion protein gene cassette by PCR amplification of MODHv1a/GNA using a 3' primer coding for six histidine residues, a stop codon and an *Xba* I restriction site. The PCR product (~500bp) was digested with *Pst* I/*Xba* I and ligated into pGAPZαB vector digested with same enzymes, to create MODHv1a/GNA/His in pGAPZαB. All constructs were verified by sequencing prior to yeast transformation.

To enable the insertion of multiple fusion protein cassettes into the yeast genome, the pGAPZαB vector was modified to contain a *Hind* III site in the GAP promoter region by site directed mutagenesis (as *Bln* I could not be used for linearisation of vector prior to yeast transformation). Primers were designed to introduce *Hind* III site 35 bp from the *Bln* I site in the GAP promoter region, Forward (5' CATTACGTTGCGGGTAAAACGG) and reverse (5' CTGGGAAGAAGCTTGCTGCAAG), where AATGCT sequence from original GAP region was changed to AAGCTT. PCR was performed using MODHv1a/GNA/His pGAPZαB vector template DNA and Phusion polymerase. The PCR product (~3.6kb) was phosphorylated using T4 polynucleotide kinase kit as per manufacturer's instructions, gel purified, and the DNA was then re-ligated using T4 DNA ligase. Re-ligated vector was transformed into *E.coli* (TOP10) cells and plasmid DNA was sequenced (using primers; Forward- 5' GTAGAAATGTCTTGGTGTCC and

Reverse- 5' AGTCTTTGGGTCAGGAGAAA) to verify the presence of a *Hind* III site in GAP region.

Transformation of single and multi copy expression cassettes

Single copy, 3 copy, 5 copy, 7 copy and 11 copy expression construct plasmids were assembled using standard molecular biology methods, as described below.

Each expression cassette consisted of pGAP region, alpha factor secretary signal, MODHv1a/GNA, six residue His tag followed by an AOX1 transcription terminator (Fig. 2a). This cassette was restricted from the backbone vector using *Bam* HI and *Bgl* II. The vector into which this cassette was to be introduced, already containing an expression cassette, was linearised with *Bam* HI. After ligation of the expression cassette, the correct resulting circular plasmid contains the unchanged Bgl II site, an intact reformed BamH I site, and ~~The cassette was then inserted utilizing the uniqueness of *Bam* HI site~~a hybrid BamH I / Bgl II site (*Bam* HI and *Bgl* II produce compatible overhangs) between the original and introduced expression cassettes, ~~and the site is modified after ligation such that it is recognized~~which cannot be digested by ~~n~~either enzyme~~s~~). It can then be linearised with BamH I to introduce further ~~Additional~~ expression cassettes ~~were introduced using the same methods~~. Each time ~~the plasmids were confirmed for the insertion of more~~further cassettes were inserted, the recombinant plasmid was checked either by releasing the complete set of cassettes (using *Bam* HI and *Bgl* II) or linearising the whole plasmid (using *Bam* HI). An overview of the preparation of multiple cassettes is shown in Fig. 2b.

Plasmid DNA (5µg) was linearised by treatment with *Hind* III, and transformed into *Pichia pastoris* strains SMD1168H and X-33 using standard Invitrogen kit protocols. Transformants were selected on antibiotic containing plates (100 µg/ml zeocin).

Transgene copy number determination by quantitative PCR

Selected yeast clones were grown in 10ml YPG medium (1% w/v yeast extract; 2% w/v peptone; 4% v/v glycerol; 100 µg/ml zeocin) in baffled flasks with shaking at 30°C for 72 h. Genomic DNA (gDNA) was extracted as described by Looke *et al.* 2011[15]. and quantified by absorbance at 260nm using a Nanodrop spectrophotometer. Primers were designed to amplify partial sequences of MODHv1a/GNA, and the *P. pastoris* actin gene (PAS_chr3_1169, encoding Uniprot protein Q9P4D1) was used as an endogenous control for gene copy number, as follows:

For MODHv1a/GNA amplification: Fwd 5'TGGTCTCTCCCGTAGCTGCTT; Rev 5'ATCGAACAAACCGATTTGGG

For Actin amplification: Fwd 5' CGGTATGTGTAAGGCCGGATA; Rev 5'ACGACCGATGGGAACACTGT

50ng of gDNA was used as a template for individual quantitative PCRs, with each reaction run in triplicate. Amplification was measured using fluorescence of SYBR Green on an Applied Biosystems StepOne instrument; Step One software was used to compare samples using the ΔCt method.

Small-scale screening for fusion protein expression

Selected clones were grown in small-scale (10 ml) cultures in YPG medium with shaking at 30° C for 48 h. After centrifugation (3000*g*, 10 min), 10µl supernatant samples were analysed by SDS-PAGE followed by western blotting using anti-GNA antibodies, as described previously ([16] Fitches & Gatehouse, 1997), except that chemiluminescence detection was carried out using coumaric acid (0.2 mM) and luminol (1.25 mM) in 1 M Tris (pH 8.5) with the addition of 0.009% (v/v) hydrogen peroxide. Recombinant snowdrop lectin (GNA), purified to 100 % homogeneity was used as a standard to allow quantitative estimation of fusion protein expression.

Bench-top fermentation experiments

All runs were carried out in a 5 l capacity BioFlo 110 (New Brunswick) fermentation vessel with 2.5 l of basal salt minimal media (MM) supplemented with PTM1 salts ([17] Cino, 1999). Fermenters were seeded with 180ml inoculum culture, grown in YPG medium (shaking at 30° C) for 48 h, and run at 30 °C. A sterile glycerol (50% v/v) feed of 1.25 l was introduced over a fermentation period of 72 h, maintaining dissolved oxygen at 30% and pH at 4.7- 4.9.

Yield estimates following fermentation, were obtained by SDS-PAGE analysis and western blotting. For western analysis samples were diluted in distilled water prior to loading. Gels (17.5 % acrylamide) were stained for total proteins with Coomassie Blue and blots were probed with anti-GNA antibodies, as described previously. Different loadings of purified GNA were used as a standard to give semi-quantitative estimates of fusion protein content.

Purification of Hv1a/GNA and MODHv1a/GNA/His fusion proteins

Purification of the non-modified recombinant Hv1a/GNA fusion protein was carried by hydrophobic interaction chromatography followed by a gel filtration clean-up step as described previously ([7] Fitches *et al.*, 2012). Purification of MOD/Hv1a/GNA/His was carried out in a single step using nickel affinity chromatography on 5 ml HisTrap crude nickel columns (GE Healthcare). Briefly, one-third volume 4x binding buffer (BB; 0.4 M sodium chloride, 0.04 M sodium acetate; pH 7.4) was added to culture supernatants, ~~and~~ which were ~~loaded~~ then loaded onto columns equilibrated with BB at a flow rate of 2- 4 ml/min. Bound protein was eluted by the addition of imidazole (200mM) in BB. The eluted protein

peak was diluted 50:50 with distilled water, dialysed against distilled water and freeze-dried. For analysis of fusion protein content a known quantity was weighed and re-suspended in distilled water at a concentration of 10 mg/ml. Following centrifugation at 12000*g* for 2 min aliquots of supernatant were run on SDS-PAGE, using GNA standards run on the same gel to allow semi-quantitative estimation of fusion protein content after staining.

Analysis of biological activity: injection bioassays

Biological activity was assessed by injection of purified Hv1a/GNA and MODHv1a/GNA/His into newly moulted fifth instar cabbage moth (*Mamestra brassicae*) larvae. Proteins were re-suspended in distilled water at concentrations of 1 – 4 µg/µl. Larvae were anesthetized with carbon dioxide for 10-20 sec prior to injecting 5 µl volumes using a Hamilton syringe with a 24 gauge needle. Controls were injected with distilled water. For each sample tested 20 larvae were injected per dose. Survival was monitored daily for 5 days.

Results and Discussion

The expression vector used as the basis for constructs was pGAPZαB, a shuttle vector propagated in *E. coli*. ~~The recombinant insecticidal fusion protein Hv1a/GNA, previously shown to be orally active against lepidopteran larvae, has potential for development as a biopesticide ([7] Fitches et. al., 2012).~~ The basic construct for heterologous expression of the Hv1a/GNA fusion protein ~~contained a coding sequence encoding~~encodes a hybrid protein composed of the yeast (*S. cerevisiae*) α-factor prepro-sequence, the ω-ACTX-Hv1A atracotoxin peptide (Hv1a), joined by a 3-alanine linker region to the coding sequence for GNA (*Galanthus nivalis* agglutinin; snowdrop lectin) polypeptide (Fig. 1a). Expression of this construct results in secretion of the insecticidal Hv1a/GNA fusion protein, into culture supernatant of transformed *P. pastoris*. To date, the production of gram quantities of intact fusion protein required for field testing, has been hindered by problems associated with proteolytic cleavage during expression by *P. pastoris*, and by the relatively low levels of expression observed for protease deficient strains carrying single copy Hv1a/GNA expression cassettes. Here we report work carried out to improve levels of expression of intact Hv1a/GNA through modification of the Hv1a/GNA construct and subsequently the generation of a modified expression vector backbone that enabled the integration of multiple Hv1a/GNA expression cassettes.

Construct modification and expression analyses

The C-terminus of the Hv1a peptide (residues 33–36) includes the sequence -VKRC-, which is similar to the sequence -EKRE- that is present in the α-factor signal

sequence of the pGAP expression vector. -EKRE- is cleaved between R and E by the *P. pastoris* Kex2 gene product. The 34th lysine residue in Hv1a was replaced with a glutamine residue by site directed mutagenesis and, following sequence verification, reassembled into an expression vector that was transformed into *P. pastoris* wild type X33 and protease deficient SMD1168H strains. Western blot analysis using anti-GNA antibodies enabled comparative analysis of the stability of the original Hv1a/GNA fusion protein with the modified MODHv1a/GNA/His in both nutrient rich (YPG) and minimal media (MM). As reported previously ([7] Fitches *et al.*, 2012) the presence of two GNA immunoreactive bands on western blots corresponds to intact fusion protein and GNA from which the Hv1a toxin has been cleaved. Due to the presence of a histidine tag intact MODHv1a/GNA/His and cleaved GNA/His migrate at higher molecular masses than the non-modified Hv1a/GNA and cleaved GNA proteins. A composite of these experiments is presented Fig. 1b&c.

Levels of intact Hv1a/GNA were found to be higher in both nutrient rich (YPG) and minimal media (MM) when the protein is expressed in the protease deficient strain as compared to the wild type strain (Fig. 1b). This has been reported previously for a fusion protein encoding a venom peptide SFl1, from the spider *Segestria florentina*, linked to GNA ([5] Fitches *et al.*, 2004). In the case of SFI1/GNA, expression by X33 cells was found to produce a 1:1 ratio of intact to cleaved protein, whereas predominantly intact SFI1/GNA was obtained using the protease deficient strain SMD1168H. Similarly for Hv1a/GNA, expression using SMD1168H cells resulted in 75-100% intact fusion protein (0-25% cleaved GNA) as compared to 50-60% in X33 (50-40 % cleaved GNA), depending on the composition of the culture media.

Modification of the C-terminus of the toxin sequence resulted in a small increase in the levels of intact fusion protein in SMD1168H cells as compared to the original construct (Fig. 1b & c). This indicated that proteolysis during processing of the protein in the golgi apparatus by Kex2 membrane bound proteases may be slightly reduced. More significantly, toxin sequence modification resulted in a marked improvement (approx. 30 % Fig. 1b & c) in the levels of intact fusion protein being expressed in the wild type X33 strain as compared to Hv1a/GNA. This is indicative of reduced susceptibility to proteolytic cleavage by extracellular yeast proteases, such as carboxypeptidase Y and proteinase B, that are activated by vacuolar Peptidase A. Reduced proteolysis of the modified fusion protein was most evident for cells grown in bench-top fermentation in minimal media (MM) as compared to small-scale shake flask cultures in nutrient rich media. Previous studies have shown that intracellular vacuolar proteases are released from yeast cells during high-density cell growth and can play a significant factor in protein degradation ([18,19] Sinha *et al.*, 2004; Li *et al.*, 2007). Thus it is likely that the modified fusion protein has improved resistance to both extracellular and intracellular yeast proteases as compared to Hv1a/GNA.

The release of intracellular proteases during high cell density fermentation may also account for the increased proteolysis of both Hv1a/GNA and MODHv1a/GNA observed in fermented cultures grown in minimal basal salt media as compared to

small-scale shake flask cultures grown in nutrient rich YPG media (Fig. 1b&c). It is also possible that the ~~presence of~~ additional proteins in YPG media may ~~increase~~ act as substrates for released proteases, decreasing the availability of the fusion proteins as substrates,~~availability~~ and ~~may~~ thus ~~have~~ contributinged to a reduction in susceptibility of the recombinant fusion proteins to proteolysis.

Analysis of biological activity: injection bioassays

Purified Hv1a/GNA and MODHv1a/GNA/His were injected into lepidopteran larvae to investigate if modification of the C-terminus of the toxin peptide had any impact upon the insecticidal activity of the fusion protein. As shown in Table 1 levels of mortality observed for *M. brassicae* larvae injected with different doses of the original or modified fusion protein were highly similar verifying that mutagenesis of the Hv1a peptide had not disrupted or altered biological activity. Hv1a is a member of the ω-ACTX-1 family of 36-37 insecticidal residue peptides isolated from the Australian funnel web spider that block insect but not vertebrate, voltage-gated calcium channels ([20] Fletcher et al., 1997). Previous alanine scanning mutagenesis studies by Tedford *et al.*, [21] 2004 have identified 3 key functional residues (Pro10, Asn27, and Arg35) that determine specific binding to insect calcium channels. Here replacement of the 34th lysine residue in Hv1a with a glutamine residue by site directed mutagenesis was selected as glutamine is known to be present at this position of other members of the ω-ACTX-1 family ([22] Tedford *et al* 2004) and was thus unlikely to disrupt biological function of the recombinant toxin.

Assembly of expression vector constructs containing multiple gene copies

The ~~expression vector used as the basis for constructs was~~ pGAPZαB~~, a~~ expression ~~shuttle~~ vector ~~propagated in *E. coli* which~~ integrates into the *P. pastoris* genome by homologous recombination at the chr2-1_0437 locus, the gene encoding glyceraldehyde-3-phosphate dehydrogenase (GAP). The plasmid contains the entire promoter region of the *GAP* gene, which both drives expression of the incorporated coding sequence in *P. pastoris*, and directs recombination at a site in the 5' UTR determined by the restriction enzyme used to linearise the plasmid prior to yeast transformation. As a result of the recombination, the entire plasmid is incorporated into the *P. pastoris* genome, and the 5' UTRs of both the endogenous *GAP* gene and the introduced gene construct are reconstructed.

To ensure multiple copies of recombinant gene carried by the vector are incorporated into the genome of *P. pastoris*, the approach followed was to produce vectors carrying multiple copies of the expression construct, using a strategy based on that described by Zhu et al., [13]. This approach is much more reliable than the earlier method of screening large numbers of transformants for "jackpot" high copy strains resulting from multiple integrations of a transforming plasmid ([23] Higgins and Cregg, 1998).

The strategy to obtain multi-copy expression vectors is summarised in Fig. 2. The vector backbone for the multi-copy plasmid was modified by insertion of a *Hind* III site near the *Bln* I site in the *GAP* gene 5' UTR; this modification changed ~~a single~~two adjacent bases, and was designed not to affect promoter function. The resulting expression vector (pGAPZαBH-MODHv1a/GNA(FP)) was transformed into *P. pastoris* and checked for expression; no difference to the original pGAPZαB expression construct was observed. This construct gave the 1-copy baseline for subsequent manipulation. A restriction fragment containing the GAP promoter, the recombinant protein coding sequence, and the *AOX* gene 3' UTR was isolated by restriction of the original expression construct with *Bam* HI and *Bgl* II, and was ligated into the original construct restricted with *Bam* HI; selection of the correct orientation of the inserted fragment gave a construct (pGAPZα-2FP) from which a fragment containing 2 copies of the GAP promoter - coding sequence - AOX 3' UTR in the same orientation ("2 copy cassette") could be isolated by restriction with *Bam* H_I and *Bgl* II, since the site where the two fragments join was not restrictable by either enzyme after the ligation. The multi-copy expression vectors were then built up by combining the 2-copy cassette with the modified expression vector pGAPZαBH-FP. For example, to produce the 3-copy expression vector, pGAPZαBH-3FP, the 2 copy cassette was ligated with pGAPZαBH-FP which had been linearised by restriction with *BamH* I. Restriction analysis of the resulting clones was used to select a recombinant where the three copies of the cassette were in the same orientation; the resulting plasmid has a single *BamH* I site, which can be used to insert further cassettes.

In practice, it was more efficient to use both 2-copy and 4-copy cassettes to produce multi-copy expression vectors; the 4-copy cassette was produced by ligation of a 2-copy cassette into the pGAPZα-2FP plasmid in the correct orientation to allow a 4-copy cassette to be excised by restriction with *BamH* I and *Bgl* II (Fig. 2). Addition of the 4-copy cassette to pGAPZαH-FP gave pGAPZαH-5FP; addition of the 2-copy and 4-copy cassettes to pGAPZαH-5FP gave pGAPZαH-7FP and pGAPZαH-9FP; addition of the 4-copy cassette to pGAPZαH-7FP gave pGAPZαH-11FP. The larger plasmids became successively more difficult to transform into *E. coli* for cloning and propagation, and production of multi-copy expression vectors with more than 11 copies of the expression construct was not found possible.

The multi-copy expression constructs pGAPZαH-3FP, pGAPZαH-5FP, pGAPZαH-7FP, pGAPZαH-9FP and pGAPZαH-11FP were verified by restriction analysis; DNA sequencing was not feasible as the multiple cassettes all contained sequences corresponding to the pGAP sequencing primers. Linearisation of the plasmids at the single *Hind* III site in the pGAPZαB backbone showed that plasmid size increased in line with predicted copy number (results not shown).

Yeast transformation and gene copy analysis of transformants

Transformation of *P. pastoris* with the multi-copy expression vectors was carried out using normal procedures; selection on high levels of zeocin was not necessary. In contrast to transformation of *E. coli* with large plasmids, no significant fall-off in numbers of transformants obtained from the larger multicopy expression vectors was observed. Expression constructs were transformed into both the "wild-type" *P. pastoris* strain X-33, and the protease deficient strain SMD1168H.

P. pastoris clones that were selected on the basis of zeocin resistance were screened for recombinant protein expression by western blot analysis of culture supernatant from small-scale shake flask cultures. These assays showed that >90% of the selected transformants produced detectable levels of the desired recombinant protein, MODHv1a/GNA/His, irrespective of the copy number of the multi-copy expression vector used for transformation. Clones were selected for further study on the basis of positive fusion protein expression in the small-scale screen.

A selection of clones produced using single copy and multi-copy expression vectors were analysed by quantitative PCR to estimate the actual copy number of the fusion protein gene(s) incorporated into the yeast genome, by comparison to an endogenous single copy gene sequence. Analysis of 8 different clones for single copy, 3-copy, 5-copy, 7 copy, 9-copy and 11 copy transformants in *P. pastoris* SMD1168H, and 11-copy transformants in *P. pastoris* X-33 is presented in Fig. 3. These results are typical of other analyses carried out on multi-copy transformants. 7/8 of the single copy transformant clones contain a single integrated gene for MODHv1a/GNA, with the remaining clone containing 2 copies. Of the nominal 3-copy clones, 4/8 do contain 3 copies, with the remaining clones containing 1-2 copies; the "5-copy" clones have 2/8 with 5 copies, 6/8 with 1-3 copies; the "7-copy" clones have 2/8 with 7 copies, 6/8 with 1-6 copies; the "9-copy" clones have 2/8 with 9 copies, and 6/8 with 2-7 copies; and the "11-copy" clones have 4/16 with 11 copies, and 12/16 with 4-8 copies. These results show that clones produced using the multi-copy expression vectors are liable to loose copies of the expression cassette during the transformation process but that intact multi-copy integration also occurs at a viable frequency. The results in the two yeast strains were similar.

Recombinant protein expression from P. pastoris clones

The small-scale screen of clones for expression of MODHv1a/GNA/His showed an approximate correspondence between levels of expression and the number of gene copies present in the yeast clones, but expression from clone to clone was variable (results not presented). This variability is likely to be due to differences in parameters such as oxygen level and pH in small-scale cultures. The influence of copy number on protein expression is not predictable [24] and to allow a more accurate comparison of production of recombinant protein, a series of clones were selected for pilot-scale bench-top fermentation, under controlled conditions. The clones were selected on the

basis of qPCR results, which showed 1, 2, 5, 9 and 11 fusion protein gene cassettes to be present in strain SMD1168H clones, and 9 and 11 copies in strain X-33 clones. Results are summarised in Table 2.

Growth parameters for the different multi-copy expression clones in the fermenter suggested that the multi-copy clones in strain X-33 grew better than any of the clones in the protease deficient strain, as would be expected. The clones in strain X-33 also had higher pellet weights at the end of the fermentation (Table 2). When culture supernatants from the fermenter runs were analysed for protein content, it was clear that the multi-copy expression clones were able to produce enhanced levels of recombinant MODHv1a/GNA. The basal level of expression from a single copy clone in strain SMD1168H for the original Hv1a/GNA construct was approx. 50 mg/l whereas for the modified construct (MODHv1a/GNA) ~~this~~ the fusion protein level was approx. 100 mg ~~FP~~ /-l culture supernatant. There were no consistent or significant differences in growth between 2, 5 and 9-copy clones in strain SMD1168H, suggesting that any increased production of recombinant protein in the multi-copy clones was not limiting growth. However, although similar yields of fusion protein (600 mg/l) were observed for the 5 and 9-copy SMD1168H clones, ~~and~~ a ~~notable~~ reduction in expression level (200 mg/l) was recorded for the 11 copy SMD1168H clone. These results suggested that the optimal copy number for fusion protein expression in the Protease deficient strain was 5. By contrast, the optimal copy number for fusion protein expression levels using wild type cells was found to be higher than that observed for protease deficient cells. Maximum expression (1 gram/l) was recorded following fermentation of a 9-copy clone. As observed for the protease deficient clones expression levels were then seen to decrease with increasing copy number.

These results are comparable with those reported by Zhu et al. (2009) [13], where production of a recombinant protein using a methanol-induced vector in *P. pastoris* increased approximately 10-fold when increasing gene copy number from 1 to 12, but decreased when vectors with higher copy numbers were introduced. In addition Zhu et al., [13] found a reduction in cell growth rate with strains carrying more than the optimal 12 copies of the transgene. In this study a reduction in cell growth rate was noted for SMD1168H cells carrying more than 5 copies of the transgene but this was not observed for X33 cells. John Is it possible that the FP5 cassette drops out of the yeast and so cells not expressing fusion protein are growing in the higher copy fermntation runs??

As shown in Fig. 4 the purified recombinant fusion protein product from multi-copy strains of *P. pastoris* gave a similar band pattern on SDS-PAGE to the product from a single copy strain. Furthermore SDS-PAGE analysis of samples derived from fermentation of the highest expressing clone demonstrated that the fusion protein is

Comment [JG2]: There is clearly recombination to decrease copy number during transformation, but once the strains are established, they should be stable. It is possible that no0n-expressing mutants are being selected for in fermentation runs, but yeast should be reasonably stable.

the most abundant protein representing more than 50% of the total protein in the culture supernatant.

Conclusion

In the present study we provide evidence to show that a single amino acid change in a recombinant hybrid protein sequence can have a significant impact on resistance to yeast proteases without loss of biological function. Furthermore, the data presented show that producing strains of *P. pastoris* containing multiple copies of a transgene expression construct results in increased expression levels of recombinant proteins when a constitutive, rather than an induced promoter is used, and identifies likely limits for recombinant protein production in this system. The increase in expression level of Hv1a/GNA has made production of this fusion protein on an industrial scale at a viable cost possible, so that it can be taken forward for evaluation as a novel biopesticide.

Acknowledgements

This work was supported by funding from the Technology Strategy Board (Project 100865).

Figure legends

Fig. 1(A) Diagrammatic representation of Hv1a/GNA and modified (MOD) Hv1a/GNA constructs showing linker region sequence and predicted molecular masses of Hv1a and GNA components. The location and identity of the modified amino acid in the Hv1a peptide is depicted in white text. **(B)** Composite of western analysis (anti-GNA antibodies) of culture supernatants derived from small-scale YPG and bench-top fermented (MM) samples of Hv1a/GNA and MODHv1a/GNA/His expressing SMD1168H or X33 clones. Arrow depicts GNA standard (25 ng). **(C)** Table showing estimated percentage of intact fusion protein present in Hv1a/GNA and MODHv1a/GNA/His supernatants derived from western analysis in 1B.

Fig. 2 Diagrammatic representation of the cloning strategy adopted to enable insertion of multiple copies of the fusion protein cassette into the *Pichia* yeast genome. **(A)** Location of Hv1a/GNA cassette and position of inserted *Hind* III site in the promoter region of pGAPZαB expression vector. **(B)** Procedure for vector restriction and ligation of fusion protein cassettes to create multi-copy expression cassettes in *E. coli* prior to transformation into yeast.

Fig. 3 Quantitative PCR analyses of yeast clones derived from transformation of single and multi-copy fusion protein cassettes. Actin was used as an endogenous control. To be completed

Fig. 4 SDS-PAGE (17.5 % acrylamide gels) analyses of **(A)** Purified Hv1a/GNA and MODHv1a/GNA/His derived from fermentation of single copy SMD1168H clones and **(B)** Culture supernatant and purified MODHv1a/GNA/His samples derived from fermentation of 9-copy X33 clone. M denotes molecular scale protein marker mix. Loading in μg of intact fusion protein or GNA standards (A & B) or μl of culture supernatant (B) is denoted.

References numbering system!

[1] R. Daly & M.T.W. Hearn, Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. J. Mol. Recog. 18(2) (2005) 119-138.

[2] R.J.M. Raemaekers, L. deMuro, J.A. Gatehouse, A.P. FordhamSkelton, Functional phytohemagglutinin (PHA) and *Galanthus nivalis* agglutinin (GNA) expressed in *Pichia pastoris* – Correct N terminal processing and secretion of heterologous proteins expressed using the PHA-E signal peptide. Eur. J. Biochem. 265 (1999) 394-403.

[3] P. Baumgartner, K. Harper, R.J. Raemaekers, A. Durieux, A.M. Gatehouse, H.V. Davies, M.A. Taylor, Large-scale production and purification of recombinant *Galanthus nivalis* agglutinin (GNA) expressed in the methylotrophic yeast *Pichia pastoris*, Biotech. Lett. 25(15) (2003) 1281-1285.

[4] L. Sulan, Z. Dongting, T. Kexuan Functional GNA expressed in E. coli with high efficiency and its effect on *Ceratovacuna lanigera* Zehntner. Appl. Microbiol. Biotech. 69 (2005) 184-191.

[5] E. Fitches, M.G. Edwards, C, Mee, E. Grishin, A.M.R. Gatehouse, J.P. Edwards, J.A. Gatehouse (2004) Fusion proteins containing insect-specific toxins as pest control agents: snowdrop lectin delivers fused insecticidal spider venom toxin to insect haemolymph following oral ingestion. J. Insect Phys. 50 (2004) 61–71.

[6] E.C. Fitches, H.A. Bell, M.E. Powell, E. Back, C. Sargiotti, R.J. Weaver, J.A. Gatehouse, Insecticidal activity of scorpion toxin (ButaIT) and snowdrop lectin (GNA) containing fusion proteins towards pest species of different orders. Pest Manag. Sci. 66 (2010) 74–83.

[7] E.C. Fitches, P. Pyati, G.F. King, J.A. Gatehouse, Fusion to Snowdrop Lectin Dramatically Enhances the Oral Activity of the Insecticidal Peptide ω-hexatoxin-Hv1a by Mediating its Delivery to the Central Nervous System. PLoS One 7(6) (2012) e39389

[8] R.E. Down, E.C. Fitches, D.P. Wiles, P. Corti, H.A. Bell J.A. Gatehouse, J.P. Edwards, Insecticidal spider venom toxin fused to snowdrop lectin is toxic to the peach-potato aphid, *Myzus persicae* (Hemiptera: Aphididae) and the rice brown planthopper, *Nilaparvata lugens* (Hemiptera:Delphacidae). Pest Manag Sci 62(2006) 77–85.

[9] N.P. Trung, E. Fitches, J.A. Gatehouse, A fusion protein containing a lepidopteran-specific toxin from the South Indian red scorpion (*Mesobuthus tamulus*) and snowdrop lectin shows oral toxicity to target insects. BMC Biotech. 6: 18 (2006).

[10] Gilleson et al 1998

[11] A. Vassileva, D. Aurora Chugh, S. Swaminathan, N. Khanna, Effect of copy number on the expression levels of Hepatitis B surface antigen in the methylotrophic yeast *Pichia pastoris.* Protein Express. & Purif. 21 (2001) 71-80.

[12] M. Mansur, C. Cabello, L. Hernandez, J. Pais, L. Varas, J. Valde, Y. Terrero, A. Hidalgo, L. Plana, V. Besada, L. Garcia, E. Lamazares, L. Castellanos, E. Martınez, Multiple gene copy number enhances insulin precursor secretion in the yeast *Pichia pastoris* Biotech. Letts. 27 (2005) 339–345.

[13] T. Zhu, M. Guo, Z. Tang, M. Zhang, Y. Zhuang, J. Chu and S. Zhang, Efficient generation of multi-copy strains for optimizing secretory expression of porcine insulin precursor in yeast *Pichia pastoris*. J. Appl. Micro. 107 (2009) 954–963.

[14] Sambrook J, Russell DW Molecular cloning: A laboratory manual, 3rd edn, Spring Harbor, New York: Cold Spring Harbor Laboratory. (2001).

[15] M. Lõoke, K. Kristjuhan, A. Kristjuhan, Extraction of genomic DNA from yeasts for PCR based applications. Biotechniques 50 (2011) 325-328.

[16] E. Fitches, J.A. Gatehouse, A comparison of the short and long term effects of insecticidal lectins on the activities of soluble and brush border enzymes of tomato moth larvae (*Lacanobia oleracea*). J. Insect Phys. 44 (1998) 1213–1224.

[17] J. Cino, High-yield protein production from *Pichia pastoris* yeast: A protocol for benchtop fermentation. Application Note. New Brunswick Scientific, (1999).

[18] J. ~~Sinha~~ Sinha ~~J~~, B.A. Plantz ~~BA~~, M. Inan M, M.M. Meagher,~~r MM.~~ Causes of proteolytic degradation of secreted recombinant proteins produced in methylotrophic yeast Pichia pastoris: Case study with recombinant ovine interferon-tau. Biotechnol. Bioeng. 89 (2005)~~;~~ ~~89:~~ 102–112. ~~Find full reference~~

[19] P. Li, A. Anumanthan, X-Gong Gao, K. Ilangovan, V.V. Suzara, N. Düzgüneş V. Renugopalakrishnan, Expression of Recombinant Proteins in *Pichia Pastoris* Appl. Biochem. Biotech. 142 (2007) 105-124.

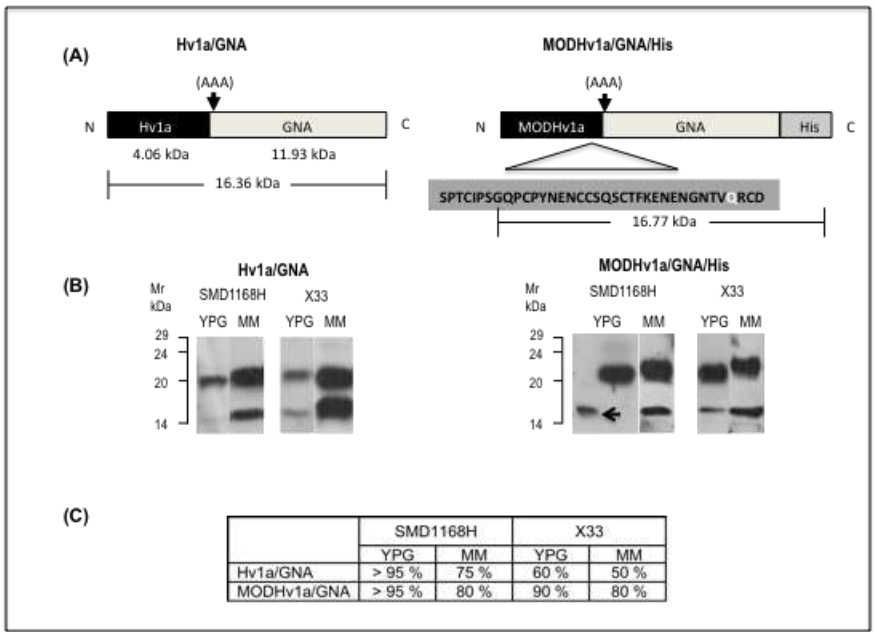
[20] J.I. Fletcher, R. Smith, S.I. O'Donoghue, M. Nilges, Connor M, M.E.H. Howden, M.J. Christie, G.F. King, The structure of a novel insecticidal neurotoxin, ω-atracotoxin-HV1, from the venom of an Australian funnel web spider. Nat Struct. Biol 4 (1997) 559–566.

[21] H.W. Tedford, N. Gilles, A. Ménez, C.J. Doering, G.W. Zamponi, G.F. King, Scanning mutagenesis of ω-atracotoxin-Hv1a reveals a spatially restricted epitope that confers selective activity against invertebrate calcium channels. J Biol Chem 279 (2004a) 44133–44140.

[22] H.W. Tedford, B.L. Sollod, F. Maggio, G.F. King, Australian funnel-web spiders: master insecticide chemists. Toxicon 43 (2004b) 601–618.

[23] D.R. Higgins, J.M. Cregg, *Pichia* Protocols. Totowa, NJ, Humana Press, 1998.

[24] K. Sreekrishna, *Pichia* Optimization of Protein Expression, Encylopedia of Industrial Biotechnology: Bioprocess, Bioseparation, and Cell Technology; ed. M.C. Flickinger, John Wiley & Sons, Inc. (2012)



	SMD1168H		X33	
	YPG	MM	YPG	MM
Hv1a/GNA	> 95 %	75 %	60 %	50 %
MODHv1a/GNA	> 95 %	80 %	90 %	80 %

Fig. 1

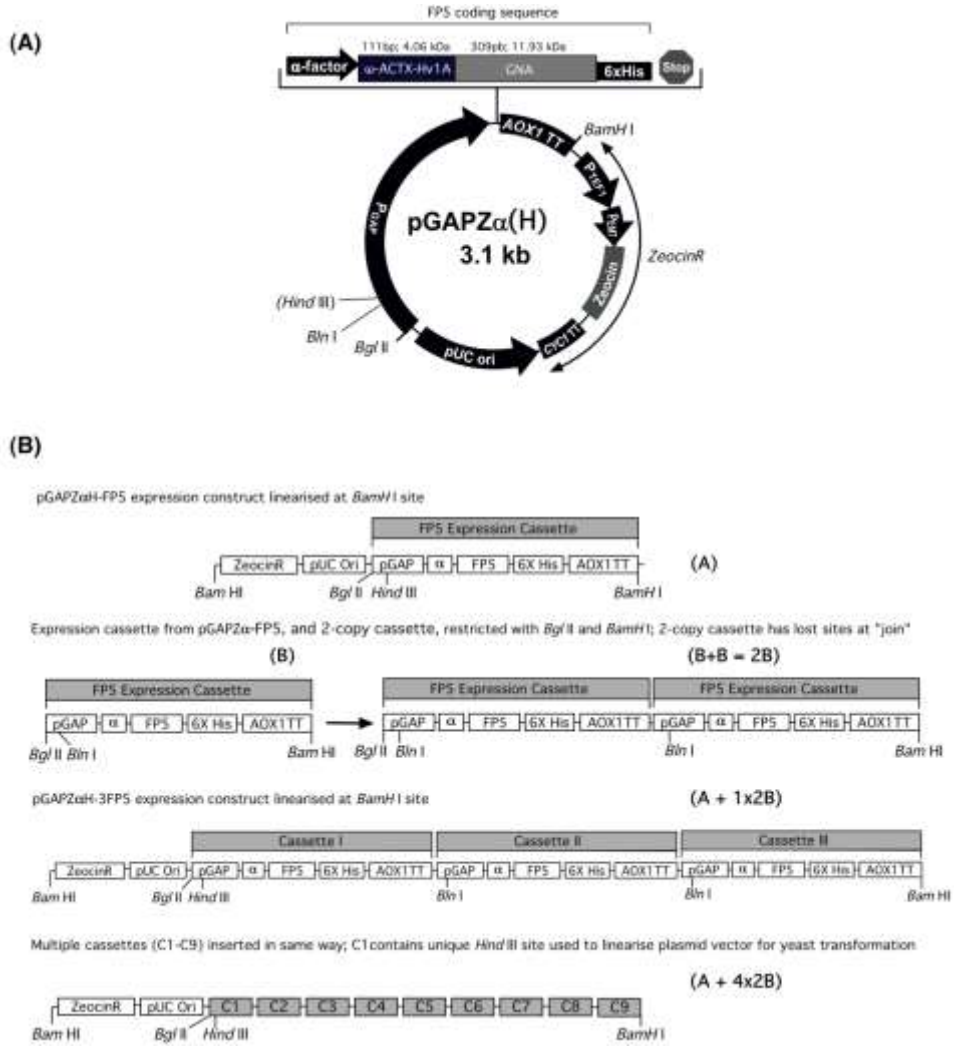


Fig. 2.We need to get rid of FP5 and replace with FP

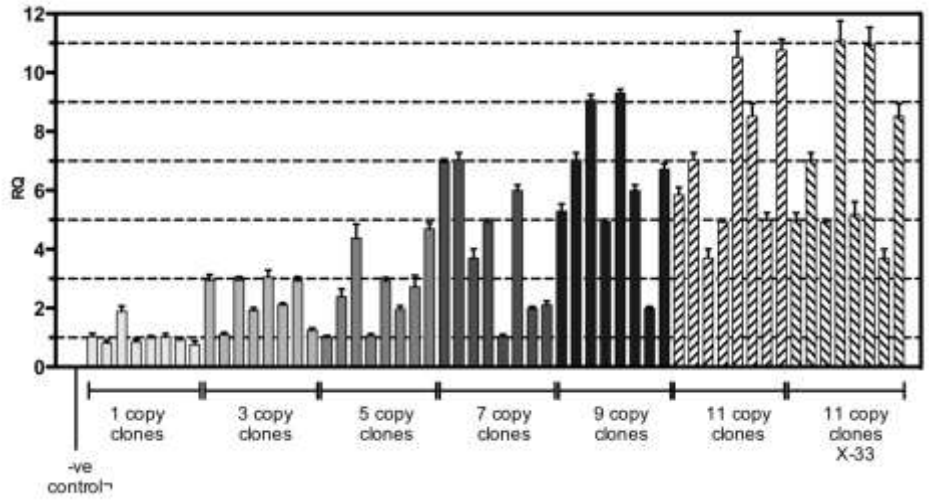

12
10
8
6
4
2
0
RQ
1 copy clones
3 copy clones
5 copy clones
7 copy clones
9 copy clones
11 copy clones
11 copy clones X-33
-ve control

Fig. 3

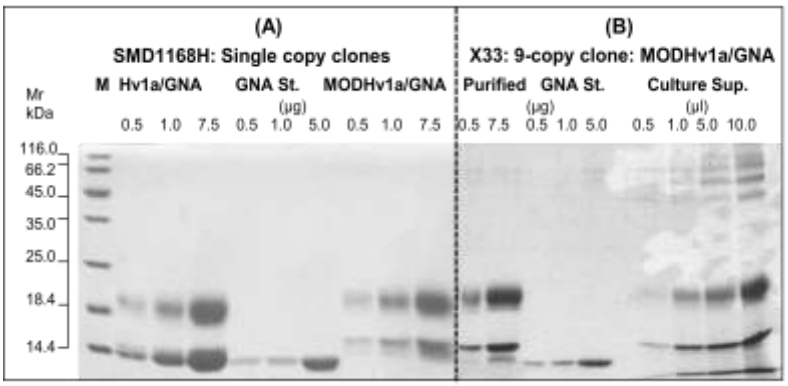

(A)
SMD1168H: Single copy clones
M Hv1a/GNA
GNA St.
MODHv1a/GNA
(µg)
0.5 1.0 7.5 0.5 1.0 5.0 0.5 1.0 7.5
(B)
X33: 9-copy clone: MODHv1a/GNA
Purified
GNA St.
Culture Sup.
(µg)
(µl)
0.5 7.5 0.5 1.0 5.0 0.5 1.0 5.0 10.0
Mr
kDa
116.0
66.2
45.0
35.0
25.0
18.4
14.4

Fig 4.

Table 1. Yields of recombinant protein product FP5 from 5 l scale fermentations of different clones of *P. pastoris* estimated by SDS-PAGE and western analysis (anti-GNA antibodies).

Clone Number	Gene Copy Number	Strain	Final wet cell density (mg/ml)	Yield estimate (mg/l)
4/1	1	SMD1168H	325	100
3/1	2	SMD1168H	316	200
3/11	2	SMD1168H	375	200
5/5	5	SMD1168H	386	600
3/7	5	SMD1168H	455	600
2/5	9	SMD1168H	350	600
7/11	11	SMD1168H	365	200
1/11	9	X-33	420	1000
5/5	11	X-33	416	600

Table 2. Mortality recorded for fifth stadium *M. brassicae* larvae 72 h after injection of different concentrations of recombinant Hv1a/GNA and MODHv1a/GNA/His re-suspended in dstilled water. Controls were injected with distilled water.

Treatment	Dose fusion protein (μg)	Mortality	Sample No.
Control	-	0	40
Hv1a/GNA	20	90	20
	10	45	20
	5	0	20
MODHv1a/GNA	20	100	20
	10	35	20
	5	0	20